CS 134: Networks Section 3
Prof. Yaron Singer Spring 2017

Note: some of the content of these section notes is borrowed from `http://arxiv.org/pdf/0706.1062v2.pdf`, which is also a great place to explore the content of this section in more depth.

1 Overview: Fitting the Parameter of a Power Law

Recall from lecture that a non-negative random variable X has a power law degree distribution if there exist constants $c, \alpha > 0$ such that $\Pr(X \geq x) = c \cdot x^{-\alpha}$ where α is called the scaling parameter. In the case where X is a discrete random variable, then there exists constant $c, \alpha > 0$ such that $\Pr(X = x) = c \cdot x^{-\alpha}$. The goal of today's section is to compute an estimate $\hat{\alpha}$ of α from some empirical data. That is,

Given random realizations of a random variable X that has a power law degree distribution with scaling parameter α, how can we estimate α?

2 The Normalizing Factor c

A power-law distribution for a continuous random variable X has density function $p(x)$ where

$$p(x)\,\mathrm{d}x = \Pr(x \leq X \leq x + \mathrm{d}x) = cx^{-\alpha}\,\mathrm{d}x.$$

We wish to find the normalizing constant c. Observe that as $x \to 0$, $p(x) \to \infty$, so we need a minimum value x_{min} for the domain of X. Since $p(x)$ is a probability density function,

$$\begin{aligned} 1 &= \int_{x=x_{min}}^{\infty} cx^{-\alpha}\,\mathrm{d}x \\ &= \frac{c}{-\alpha+1} \cdot x^{-\alpha+1}\big|_{x_{min}}^{\infty} \\ &= \frac{c}{-\alpha+1} \cdot \left(-{x_{min}}^{-\alpha+1}\right) \end{aligned}$$

and

$$c = \frac{\alpha - 1}{{x_{min}}^{-\alpha+1}}.$$

We therefore get that the density function of a continuous random variable is

$$p(x) = \frac{\alpha-1}{x_{min}}\left(\frac{x}{x_{min}}\right)^{-\alpha} \tag{1}$$

for $\alpha > 1$.

3 Fitting via a Log-Log Plot

By taking the logarithm on both sides of Equation (1), we get

$$\log p(x) = -\alpha \log(x) + \alpha \log(x_{min}) + \log\left(\frac{\alpha - 1}{x_{min}}\right). \tag{2}$$

Observe that Equation (2) implies that the power-law distribution is a straight line on a log-log plot, i.e., on a log-scale on both axis. A first technique to estimate α is the following:

- Plot the empirical data on a log-log plot.
- If the data forms a straight line, this indicates that the data may follow a power-law distribution.
- To estimate the scaling parameter α, measure the slope $\hat{\alpha}$ of the straight line.

3.1 Ordinary Least Squares Regression

We'll briefly discuss a simple method for fitting a linear function. If you're interested in this material, take CS 181 or some STAT class.

We have a set of points $\{(x_i, y_i)\}_{i=1}^{N}$, with x_i a d-dimensional column vector and y_i a scalar. We want to find a d-dimensional column vector w of weights such that for all the (x_i, y_i), $y_i = x_i^\top w$. Of course, we likely will not be able to find w such that all the pairs have exact inequality, so we define an loss function that tells us how bad our model is. We use the sum of square residuals, or the sum of the squared differences between our estimates and the true value:

$$L(w) = \sum_{i=1}^{N} (y_i - x_i^\top w)^2$$

Alternatively, we can write this in terms of matrices:

$$\begin{aligned} &= (Y - Xw)^\top (Y - Xw) \\ &= Y^\top Y - 2Y^\top Xw + w^\top X^\top Xw \end{aligned}$$

where X is a matrix where the rows are the x_i and Y is a column vector of the y_i. We can minimize this loss function because this function is convex, so that we only need to find where the derivative of the function is zero:

$$\begin{aligned} \frac{d}{dw} L(w) &= 0 \\ -2Y^\top X + 2X^\top Xw &= 0 \\ -2X^\top (Y - Xw) &= 0 \\ X^\top (Y - Xw) &= 0 \\ X^\top Y &= X^\top Xw \\ w &= (X^\top X)^{-1} X^\top Y \end{aligned}$$

Thus, we have a closed form expression for the optimal w in terms of our data! Unfortunately, for reasons which we will not cover in this class, this method can have large error in the estimated parameter $\hat{\alpha}$.[1]

[1] These reasons are given in Appendix A of `http://arxiv.org/pdf/0706.1062v2.pdf`

4 Fitting via the Method of Maximum Likelihood

We now look at an alternate method to estimate α, called the method of maximum likelihood. Let $x_1, \cdots, x_n$ be the points in our data. We wish to find $\hat{\alpha}$ that maximizes the probability that $x_1, \cdots, x_n$ are drawn from a power law distribution with parameter $\hat{\alpha}$. This method can be thought as finding the parameter for which the observed data is the most likely to have been drawn from that distribution.

4.1 For continuous power law distributions

The probability that these points were drawn from a continuous power law distribution with parameter α is proportional to

$$\prod_{i=1}^{n} p(x_i|\alpha) = \prod_{i=1}^{n} \frac{\alpha - 1}{x_{min}} \left(\frac{x_i}{x_{min}}\right)^{-\alpha}.$$

The value $\hat{\alpha}$ for which this quantity is maximized is called the maximum likelihood estimate (MLE) and we wish to derive it. This expression and its logarithm have the same maximum since the logarithm is a monotone increasing function, so we wish to find the maximum of

$$\begin{aligned}
\log\left(\prod_{i=1}^{n} \frac{\alpha - 1}{x_{min}} \left(\frac{x_i}{x_{min}}\right)^{-\alpha}\right) &= \sum_{i=1}^{n} \log\left(\frac{\alpha - 1}{x_{min}} \left(\frac{x_i}{x_{min}}\right)^{-\alpha}\right) \\
&= \sum_{i=1}^{n} \log\left(\frac{\alpha - 1}{x_{min}}\right) - \alpha \log\left(\frac{x_i}{x_{min}}\right) \\
&= n \log\left(\frac{\alpha - 1}{x_{min}}\right) - \alpha \sum_{i=1}^{n} \log\left(\frac{x_i}{x_{min}}\right).
\end{aligned}$$

This last expression is concave as a function of α, so its maximum is reached when its derivative is equal to 0. So we wish to find $\hat{\alpha}$ such that

$$\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}\hat{\alpha}} \left(n \log\left(\frac{\hat{\alpha} - 1}{x_{min}}\right) - \hat{\alpha} \sum_{i=1}^{n} \log\left(\frac{x_i}{x_{min}}\right)\right) \\
&= \frac{n}{x_{min}} \frac{x_{min}}{\hat{\alpha} - 1} - \sum_{i=1}^{n} \log\left(\frac{x_i}{x_{min}}\right).
\end{aligned}$$

We therefore get that $(\hat{\alpha} - 1) \sum_{i=1}^{n} \log\left(\frac{x_i}{x_{min}}\right) = n$ and

$$\hat{\alpha} = 1 + n \left(\sum_{i=1}^{n} \log \frac{x_i}{x_{min}}\right)^{-1}.$$

4.2 For integral power law distributions

In the case of *integral power law* distributions, the true integral points are approximated as continuous reals rounded to the nearest integer. We will not describe this approximation in details. The maximum likelihood estimator is approximated with

$$\hat{\alpha} \approx 1 + n \left(\sum_{i=1}^{n} \log \frac{x_i}{x_{min} - \frac{1}{2}} \right)^{-1}$$

using a similar analysis as in the continuous case.